

Available online at www.sciencedirect.com**ScienceDirect**

Procedia Technology 21 (2015) 49 – 53

Procedia
Technology

SMART GRID Technologies, August 6-8, 2015

Stream Computing: Opportunities and Challenges in Smart Grid

Shibily Joseph^a, Jasmin E A^{b*}, Soumya Chandran^c^aResearch Scholar, Govt.Engineering College ,Thrissur- 680009,Kerala, India^bAssociate Professor, Govt.Engineering College ,Thrissur- 680009,Kerala, India^cAsst. Professor, Govt.Engineering College ,Thrissur- 680009,Kerala, India

Abstract

Traditional Power Grid is transformed to Smart Grid by utilizing the technological advancements of Information and Communication Technology. In Smart Grid data is flowing between different components of the system. Advanced and online analytic on this massive data is required to trigger instantaneous action for grid operation and management. Traditional Big Data handling techniques store and analyze high volume data with variety, but failed to handle high velocity data. Stream computing can analyze high velocity data with variety which is essentially required in online data analytics. This paper compared different Big Data analysis techniques and reveals the importance of stream computing and its opportunities in Smart Grid data analytics.

© 2015 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of Amrita School of Engineering, Amrita Vishwa Vidyapeetham University

Keywords: Smart Grid; Advanced metering infrastructure; Stream computing; Data analytics; Big data; Outage detection

1. Introduction

In our daily life, many dramatic changes are being witnessed, which are mainly influenced by the advancements of information and communication technology (ICT). Power grid is not an exception for this, which is also transformed to a new grid termed as Smart Grid. Thus Smart Grid is a modernized electrical grid that uses information and communications technology to gather and act on information, such as information about the behaviours of suppliers and consumers, in an automated fashion to improve the efficiency, reliability, economics, and sustainability of the production and distribution of electricity [1].

* Corresponding author. Tel.: +91-9495465409

E-mail address: ejasmin@gmail.com

The major components of smart grid are power stations, power transmission lines, pluggable electric vehicles (PEV) charging stations, residential subdivision installed with solar panels, residential complex with advanced metering infrastructure (AMI) and energy smart houses with electric appliances connected to the smart grid. One of the major characteristic of smart grid is the seamless interaction between these components using networks. All these components generate data with high velocity, variety and finally lead to huge volume of data. Advanced data analytics leverages this data to achieve the goals of smart grid like efficiency, reliability, fault tolerance etc. Stored processing of these data will not be a feasible solution for all situations. Fraud detection, system monitoring, outage detection are some typical examples in smart grid which require on line data analytics. Among the smart grid components, penetration of smart meter deployment is very less compared to its benefits. Direct operational benefits of advanced Metering Infrastructure (AMI) implementation are meter reading automation, operational efficiencies in field and meter services, reduction in unaccounted energy, operational efficiencies in billing and customer management, improvement in capital spend efficiency and improvement in outage management efficiency.

Data storage and analysis is challenging in smart grid due to the volume, velocity and variety of data generated by different components of smart grid. This paper will give an introduction to big data followed by overview of big data analytics, big data architecture and big data tools. Another important section is the introduction to stream computing, stream computing architecture and frame works. A table comparing the features of Data warehouses (DWH), Hadoop and stream computing is presented. Literature reviews on different research works in stream computing applied in smart grid and conclusion which specify the importance of stream computing to be applied in smart grid are also included.

2. Big Data

Big data is defined as a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a “massive scale” [2]. Big data is now characterized by four V's which are volume, velocity, variety and value. Including all four V's we can say that big data extract deep value from high velocity, volume and variety data using advanced analytics. In smart grid phasor measurement units (PMU) data, smart meter data, sensor data, whether data are some examples of big data[5].

2.1 Big Data Analytics and tools

Big data analytics is the use of advanced analytical techniques against very large, diverse data sets that include different types such as structured/unstructured and sizes varying from terra bytes to zeta bytes in different processing modes (streaming/ batch) [3]. Data analysis will be performed in batch wise or in real time. Batch processing analytics commonly based on Map Reduce system. Such systems envisage the volume and variety of data. Real time systems analyze the data on the fly and perform event pattern analysis or continuous queries. Such systems process continuous unbounded stream of data. This type of processing is called stream computing, especially required for certain application areas like trading, fraud detection, system monitoring, outage detection and management, demand response programs in smart grid.

A good big data architecture combines, Data warehouses (DWH), Hadoop and real time processing. Big data architecture comprises tools for storage, complex computing, immediate reaction, and monitoring of streaming data in real time [6]. Data warehouses (DWH) are used for storage. Apache Hadoop is an open source software library, is a frame work that allows for the distributed processing of large data sets across clusters of commodity hardware using simple programming models. It is designed to scale up from single servers to thousands of machine each offering local computation and storage. This follows a transaction based architecture, where events are stored on main frame or database, analyzed and action performed. At the same time stream computing tools monitor millions of events in a specific time window to react proactively, they are behavior based architecture where events are analyzed in real time and action performed and then stored in databases for further analytics.

2.2 Stream Computing

A big data architecture contains several parts. Data warehouses stores large volume of structured data. Here importance is given for the volume only. Masses of structured and semi-structured historical data are stored in Hadoop which is stressed by volume and variety of data. Stream processing is used for fast data requirements, which traps the velocity and variety of the data. But three of these are complementary to each other [4]. Stream processing handles fast data where Hadoop and data warehouse (DWH) handles stored big data.

Stream computing can be applied on high velocity flows of data from real time sources such as market data, internet of things (IoT), Mobile, Sensors, Click Stream and even transactions. Stream computing enables organizations to analyze and act upon rapidly changing data in real time, enhance existing models with new insights, capture analyze and act on insight before opportunities are lost forever, and to move from batch processing to real time analytical decisions. Stream computing capability makes possible advances in any industry wrestling with the challenge of processing the flood of data created every day. Health care, telecommunication, utility companies, municipal transits, national security and many more industries can utilize stream computing. Stream computing system provides a window into today's data, where before we only had access to yesterdays data. Typical use cases of stream computing for energy industry and utility companies are distribution load forecasting and scheduling, create targeted customer offerings, condition based maintenance, enable customer energy management and smart meter analytics.

2.3 Stream Computing Architecture

Stream processing is designed to analyze and act on real-time streaming data, using “continuous queries” (i.e. SQL-type queries that operate over time and buffer windows). Commonly streams are implemented as tuples with well-defined structure [13]. Stream tuples are stamped with date of its occurrence. Also containing name value pair describing different attributes. Essential to stream processing is Streaming Analytics, or the ability to continuously calculate mathematical or statistical analytics on the fly within the stream (e.g. continuous queries). Windowing is a mechanism used in stream computing to perform operations such as join, sum, AVG on these moving streams. Windows can be defined physically (in terms of Time) or logically (in terms of number of elements). They have fixed or moving boundaries, leading to different types of windows (e.g. fixed window, moving window, land mark window) [13]. Stream processing solutions are designed to handle high volume in real time with a scalable, highly available and fault tolerant architecture. This enables analysis of data in motion. The data flow graph of a stream processor consists of stream source; filter, operators and stream sink as shown in Fig. 1. (a). A stream processing application is a collection of operators connected by streams.

Stream computing architecture consists of 1)Server for processing real-time streaming event data at high throughputs and low latency (usually in-memory), 2)interactive development environment IDE; which ideally offers visual development, debugging and testing of stream processing processes using streaming operators for filtering, aggregation, correlation, time windows, transformation, 3) Database Connectors, 4)Streaming analytics; a user interface, which allows monitoring, management and real-time analytics for live streaming data. Automated alerts and human reactions should also be possible. 5) Live data mart and/or Operational business intelligence [4]. Schematic diagram of a ideal stream computing frame work is shown in Fig 1. (b). Software technologies for stream processing are DBMS, rule engines and stream processing engines. Common main memory DBMS and rule engines are to be redesigned to use in stream computing. The 8 requirements of real time stream computing are given in [9].

2.4 Stream Computing Framework

As stream computing is a emerging technology, the tools and frameworks available today are not developed much. A good stream computing frame work expected to have all five components which are shown in fig 1.b. The major frame works are Apache Storm, Apache Spark, IBM Infosphere Streams and TIBCO StreamBase [4]. DWH, Hadoop and Stream Computing are major players in big data. They are complementary to each other. Table 1 shows a comparison of different capabilities of DWH, Hadoop and stream computing.

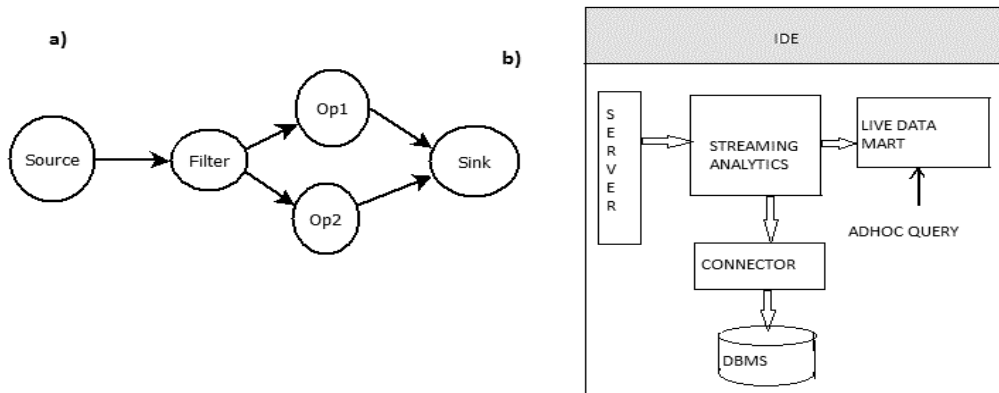


Fig. 1. a) Data flow graph of a stream processor; b) Components of stream computing framework

Table 1. Comparison of DWH, Hadoop and Stream Computing

Characteristics	Data warehouse	Hadoop	Stream computing
Type of Data Stored	Structured	Structured & Unstructured	No storage
Storage purpose	Reporting & dash board	Long running computations	Real time analytics
Age of data	Old	Past	Current/new data
Size of data	Terra/Peta bytes	Giga Bytes	Kilo Bytes
Speed of processing	Peta bytes /day	Kbps	Mbps
Implementation cost	High	Medium	Low
Volume	High	High	Low
Velocity	Nil	Nil	High
Variety	Nil	High	High

3. Related Work

In this section, some of the important works done in the smart grid application areas are reviewed where stream computing technique is used to solve the problem. Stream computing could be applied broadly by the power utilities industry to minimize network latency and function as a key component for demand response management [10]. A. Martin et.al pointed out that there is a steadily increase of data with each new installations of smart meter and the complexity of data processing task is affected by heavily fluctuations during the course of a day due to behavioral patterns[12]. This necessitates that data processing technology suitable for smart grid data must be highly scalable in order to grow organically with steadily increasing amount of data and elastic in order to consume only the required amount of resources despite the fluctuations along the day. Authors explore the combination of an elastic event stream processing(ESP) system named StreamMine3G and cloud technologies such as Amazon EC2 in the context of energy fore casing. [13] discussed the possibility to exchange data stream related to home devices and their behavior. They proposed a data stream model related to home device description and state changes. They mentioned about the existing data models in the field of electricity, which is IEC 62056 commonly known as DLMS/COSEM protocol widely used for tariff and load control [7]. This is a novel work which presents a concrete data stream model for device events. Z.Aung in [14] pointed out that database systems are one of the key stones of ICT infrastructure that provides smartness to the smart grid. Among the big software companies in data centric business, Teradata, Oracle, SAS, SAP, IBM, Microsoft and Google are active players in the smart grid area. The purpose of

data mining otherwise known as data analytics is to uncover the knowledge or interesting patterns of data that lie within a large database and use them for decision support at various levels. In D.Alahakoon and X.Yu's opinion, smart meters are the basic building block of smart grid [8]. The key functionality of the smart meter is the capture and transfer of data relating to the consumption and events such as power quality and meter status. The data has high volume, speed of collection and complexity. This paper pointed out the future needs of real time analysis, stream processing and bridging diverse data types together

4. Conclusion and Future Work

In this paper, analysis is being done on whether smart grid data come under big data. From the user scenarios, it is clear that smart grid data has the volume, velocity and variety to be qualified as big data. Then, a detailed analysis of use of big data analytics for smart grid, application areas, technologies is made followed by importance of stream computing. The different data analysis technologies are then compared. The ultimate purpose of data collection is to generate great business value for the utility companies and to the consumers using fast and efficient analytical techniques. For this, fast and efficient big data analytics on smart grid data is required. So, the possibilities of different big data techniques on different use cases of smart grid are explored. It revealed that DWH and Hadoop alone cannot perform all the necessary analytics on smart grid. Real time analysis is necessary in certain situations. This brings tremendous opportunities for stream computing in smart grid.

References

- [1] U.S. Department of Energy, Smart Grid / Department of Energy, <http://www.energy.gov> Retrieved 2015-03-10.
- [2] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, Samee Ullah Khan, The rise of big data on cloud computing: Review and open research issues, *Information Systems*, Volume 47, January 2015, p. 98-115.
- [3] <http://www-01.ibm.com/software/data/infosphere/hadoop/what-is-big-data-analytics.html>.
- [4] Kai Wahner, Real time Stream processing as Game changer in Big Data World with Hadoop and Data Warehouse,
- [5] Khan, Mukhtaj, Phillip M. Ashton, Maozhen Li, Gareth A. Taylor, Ioana Pisica, and Junyong Liu. Parallel detrended fluctuation analysis for fast event detection on massive pmu data", *IEEE Transactions on Smart Grid*, Vo.1 6, 2015
- [6] Shvachko, Konstantin, Hairong Kuang, Sanjay Radia, and Robert Chansler. The hadoop distributed file system, In *Mass Storage Systems and Technologies (MSST)*, 2010 IEEE 26th Symposium on, p. 1-10.
- [7] Huang, Wei, Cheng LI, Yu-hao CUI, Zheng XIONG, and Xin-jia LI. Research on Application of IEC 62056 Standard to Electric Power Measure [J], *Jiangsu Electrical Engineering* 1 (2010): 010.
- [8] Alahakoon, Daminda, and Xinghuo Yu. Advanced analytics for harnessing the power of smart meter big data in Intelligent Energy Systems (IWIES), 2013 IEEE International Workshop on, p. 40-45. IEEE, 2013.
- [9] Ajoy Kumar, Redefining Smart grid Architectural Thinking using Stream Computing- Cognizant
- [10] Aiello, Marco, and Giuliano Andrea Pagani. The smart grid's data generating potentials. In *Computer Science and Information Systems (FedCSIS)*, 2014 Federated Conference on, p. 9-16. IEEE, 2014.
- [11] Simmhan, Yogesh, Baohua Cao, Michail Giakkoupis, and Viktor K. Prasanna. Adaptive rate stream processing for smart grid applications on clouds." *Proceedings of the 2nd international workshop on Scientific cloud computing*, p. 33-38. ACM, 2011.
- [12] El Mahrsi, Mohamed Khalil, Sylvie Vignes, Georges Hébrail, and M-L. Picard. A data stream model for home device description In *Research Challenges in Information Science*, 2009. RCIS 2009. Third International Conference on, p. 395-402. IEEE, 2009.
- [13] Simmhan, Yogesh, Baohua Cao, Michail Giakkoupis, and Viktor K. Prasanna. Adaptive rate stream processing for smart grid applications on clouds, *Proceedings of the 2nd international workshop on Scientific cloud computing*, p. 33-38. ACM, 2011.
- [14] Aung, Zeyar. *Database Systems for the Smart Grid*, p. 151-168. Springer London, 2013.